

Minería de datos y Patrones

Version 2024-I

Fisher y Covarianzas

[Capítulo 3]

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electronica

Universidad Popular del Cesar

×	class 1
+	class 2
▽	class 3

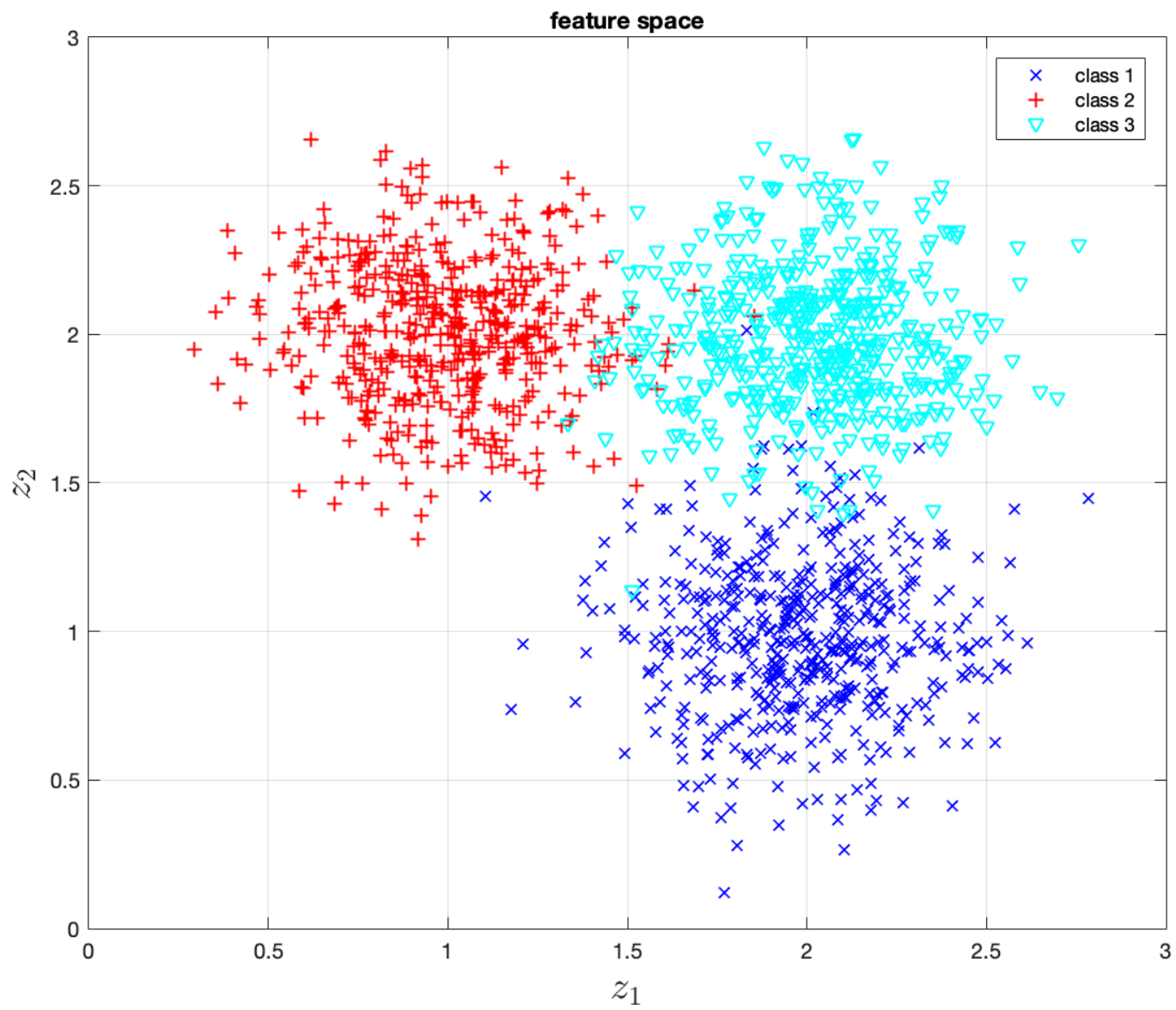
$$k = 1 \dots 3$$

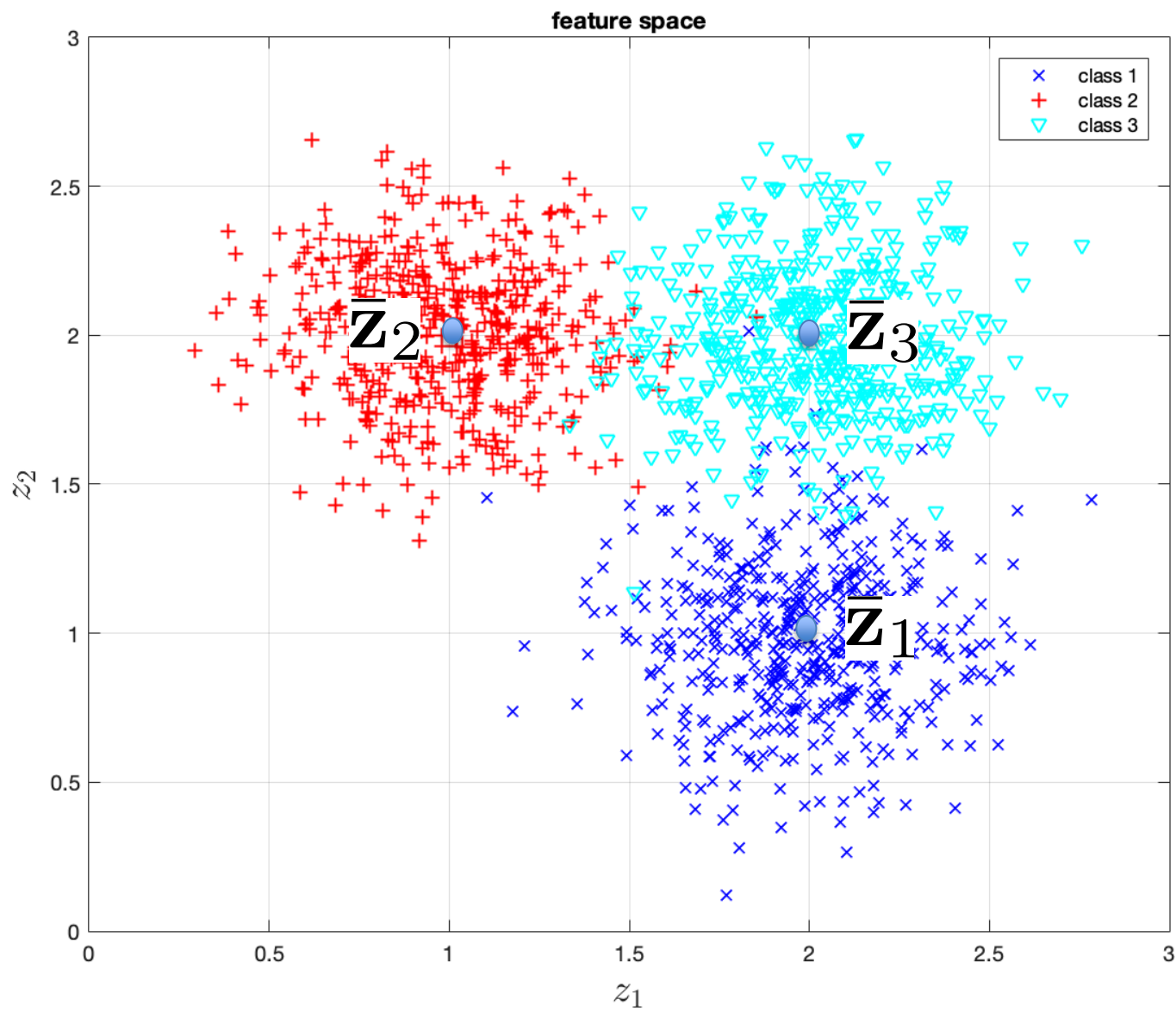


Ejemplo: 3 clases y 2 características



(z_1, z_2)

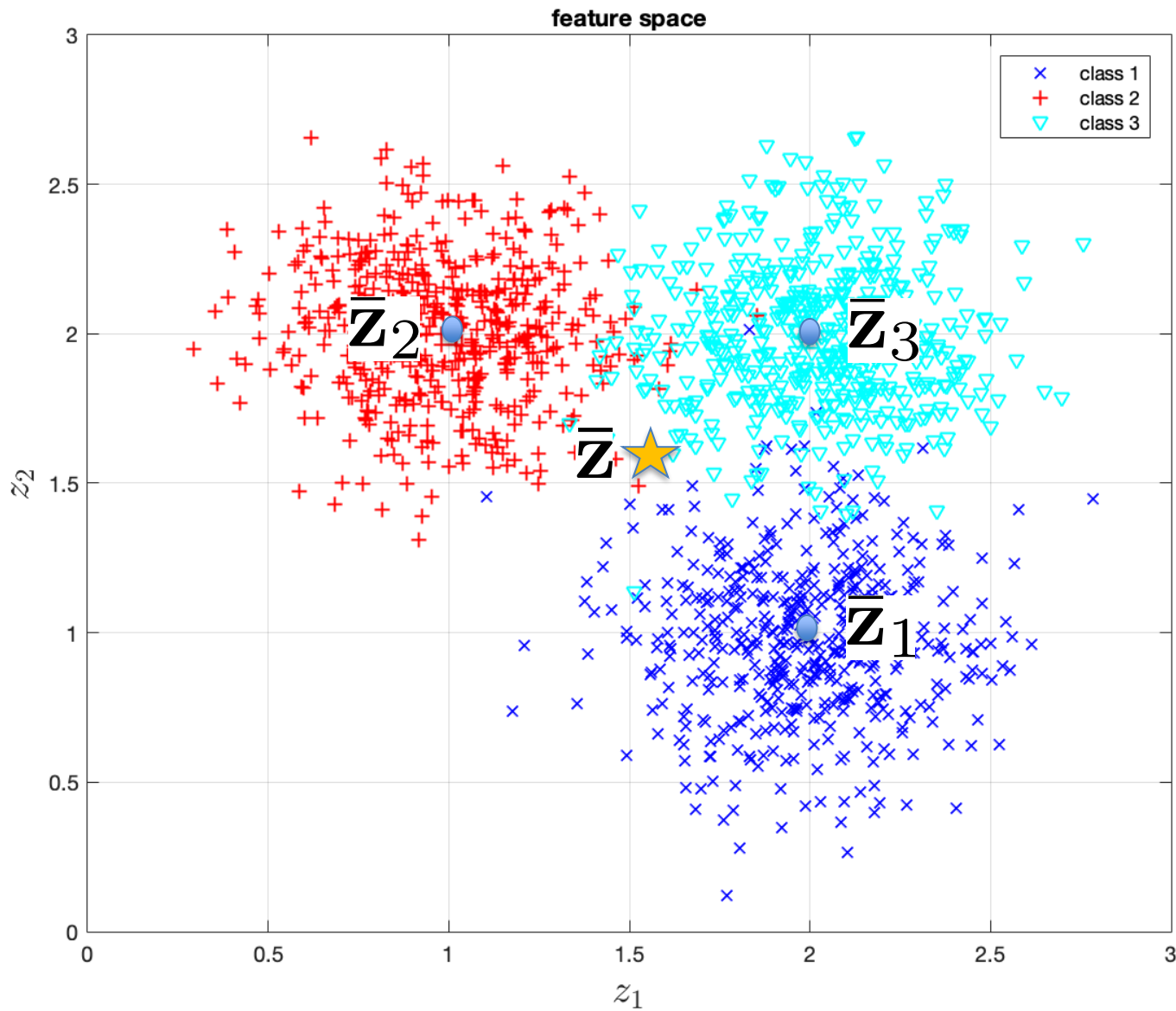




Centro de Masa
de cada clase

\bar{z}_k

$k = 1 \dots 3$



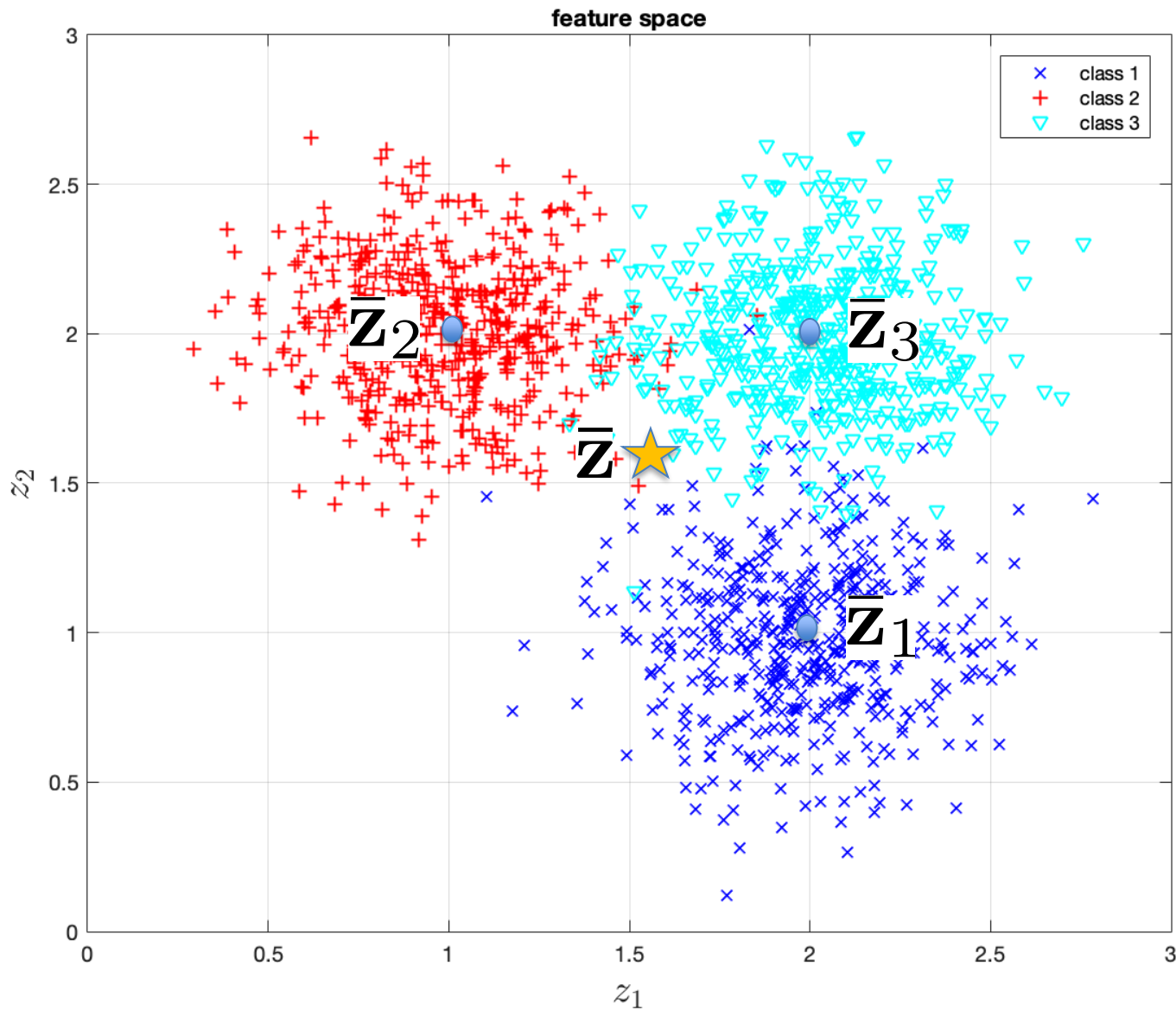
Centro de Masa
de cada clase

\bar{z}_k

$k = 1 \dots 3$

Centro de Masa
de todas las
muestras

\bar{z}



Centro de Masa
de cada clase

\bar{z}_k

$k = 1 \dots 3$

Centro de Masa
de todas las
muestras

\bar{z}

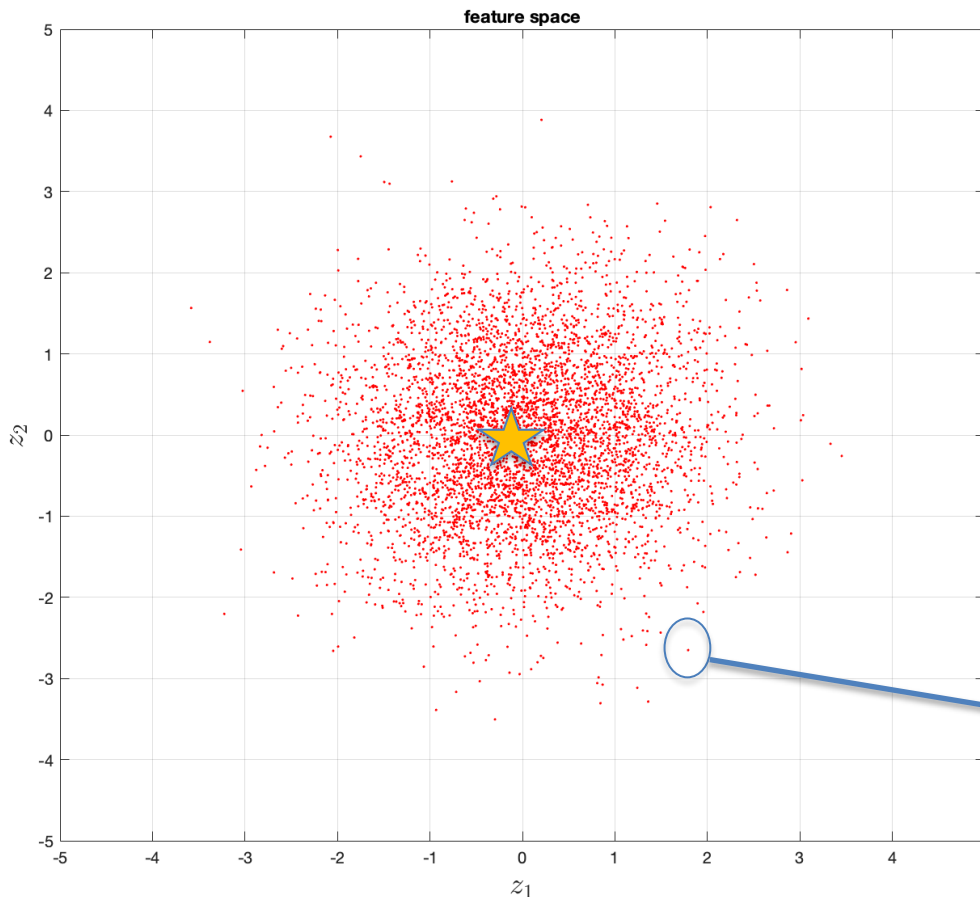
Probabilidad a
priori de que
ocurra clase k

p_k

Matriz de Covarianza

(recordatorio)

$$\mathbf{C} = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{z}_j - \bar{\mathbf{z}})(\mathbf{z}_j - \bar{\mathbf{z}})^T$$



N número de muestras

centro de masa

$$\bar{\mathbf{z}} = \frac{1}{N} \sum_{j=1}^N \mathbf{z}_j$$

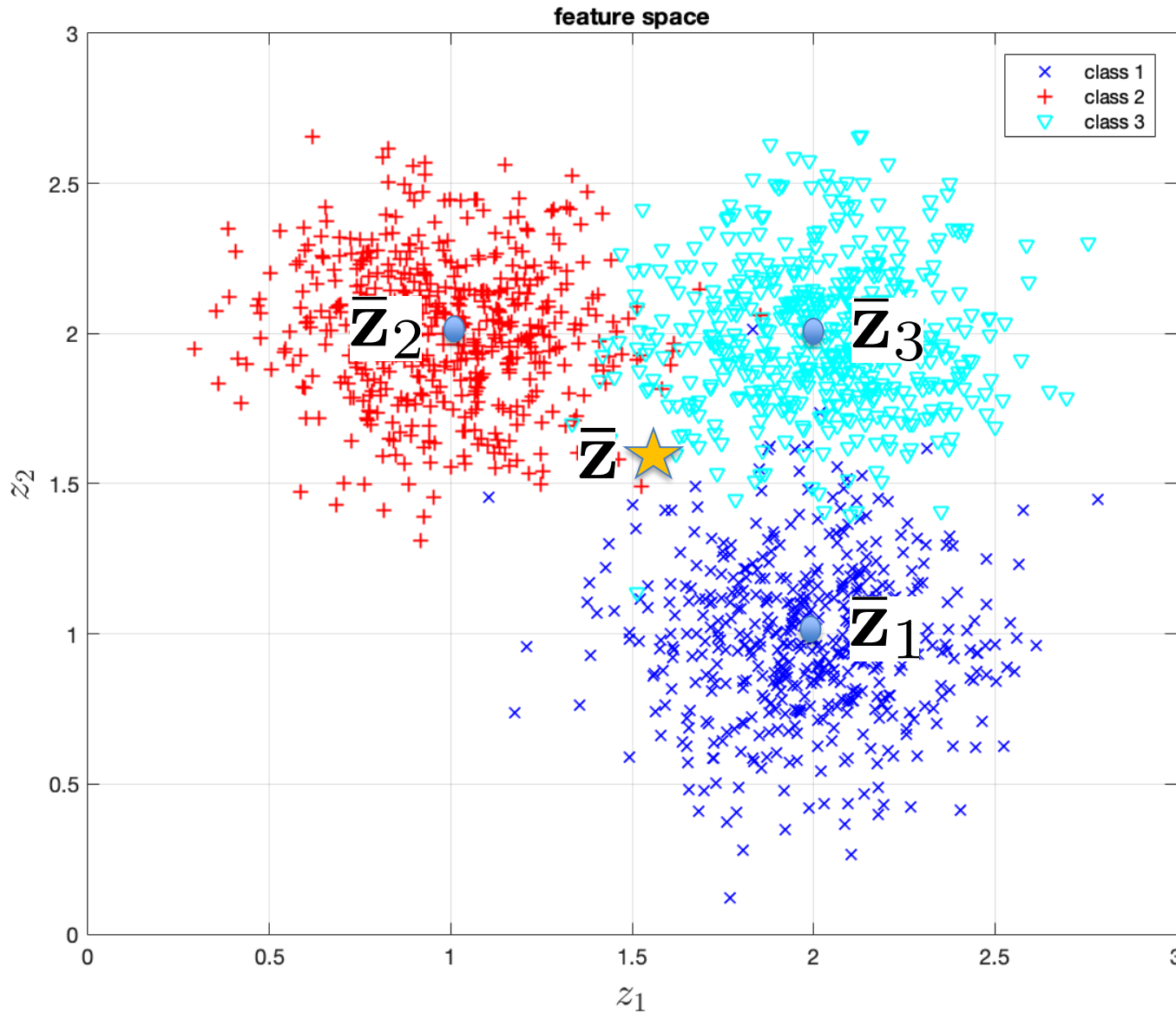


muestra j

$$\mathbf{z}_j = (z_{1j}, z_{2j})^T$$

$j = 1 \dots N$

Definiciones



Centro de Masa
de cada clase

\bar{z}_k

$k = 1 \dots 3$

Centro de Masa
de todas las
muestras

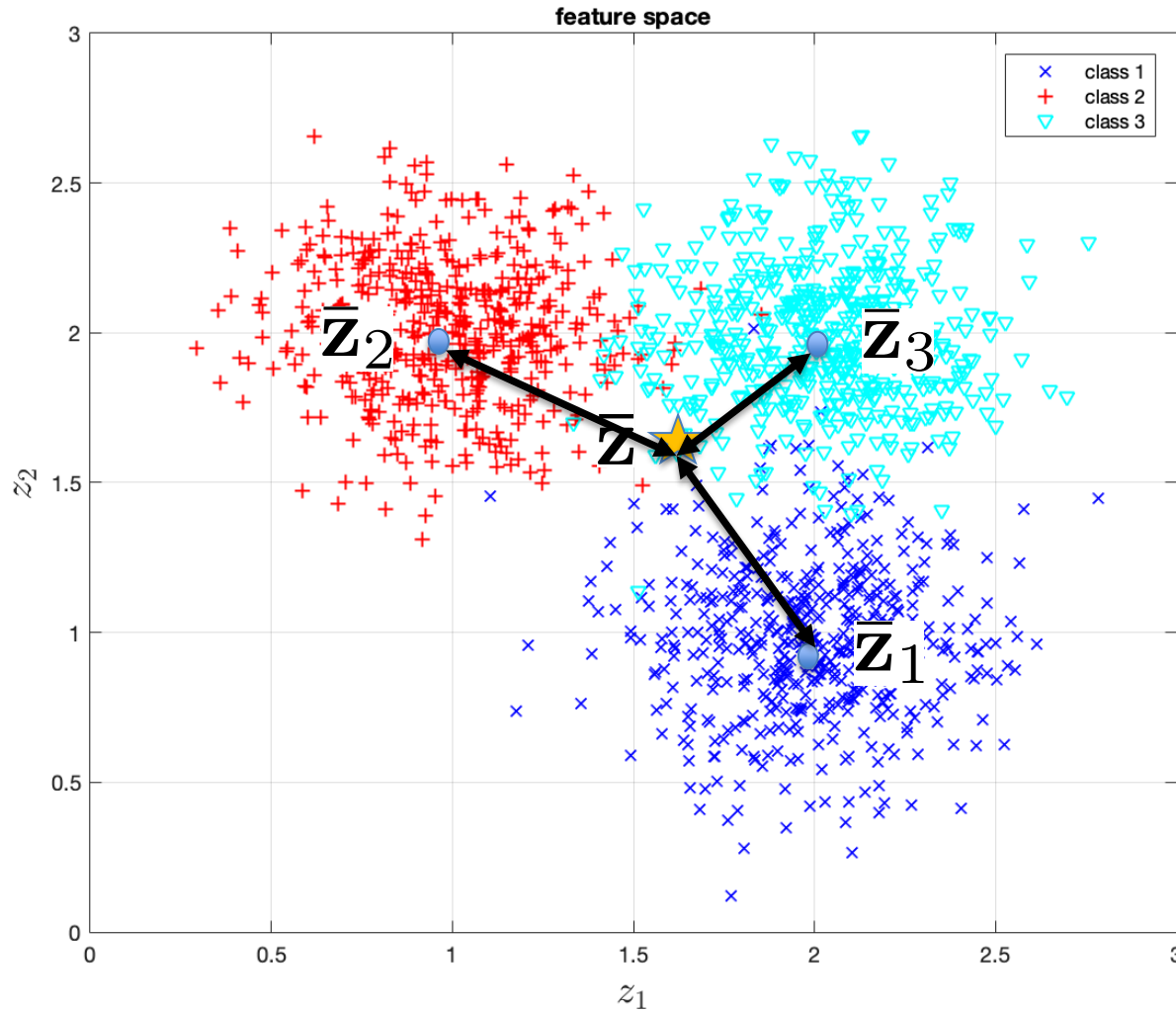
\bar{z}

Probabilidad a
priori de que
ocurra clase k

p_k

Covarianza Inter-Clase (between class covariance)

$$\mathbf{C}_b = \sum p_k (\bar{\mathbf{z}}_k - \bar{\mathbf{z}})(\bar{\mathbf{z}}_k - \bar{\mathbf{z}})^\top$$



Covarianza Intra-Clase (within-class covariance)

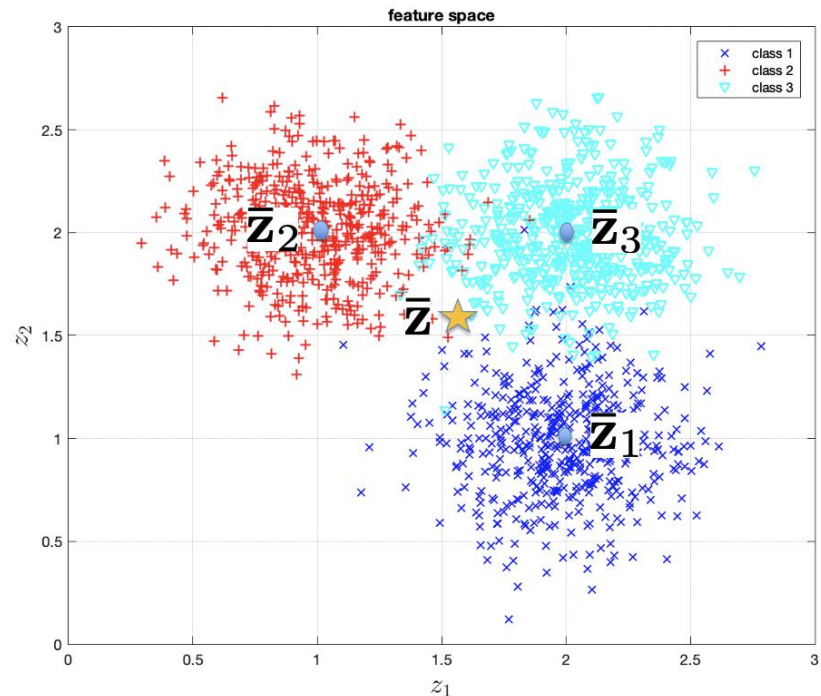
$$\mathbf{C}_w = \sum_{k=1}^K p_k \mathbf{C}_k,$$

Promedio ponderado de las Covarianzas de la cada clase

$$\mathbf{C}_k = \frac{1}{N_k - 1} \sum_{j=1}^{N_k} (\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)(\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)^T$$

Covarianza de la clase k

N_k muestras de la clase k



Covarianza Intra-Clase (within-class covariance)

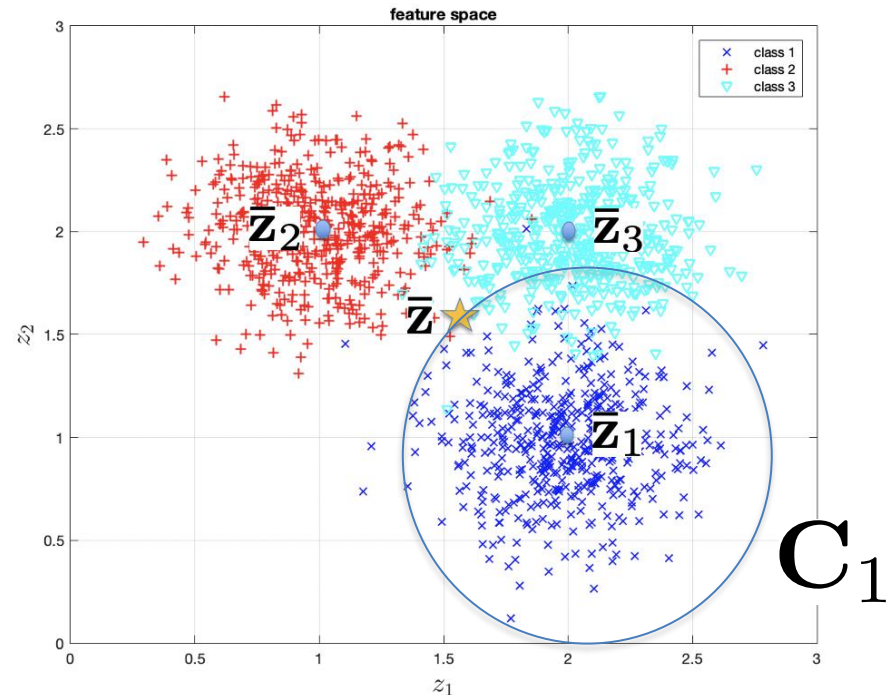
$$\mathbf{C}_w = \sum_{k=1}^K p_k \mathbf{C}_k,$$

Promedio ponderado de las Covarianzas de la cada clase

$$\mathbf{C}_k = \frac{1}{N_k - 1} \sum_{j=1}^{N_k} (\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)(\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)^T$$

Covarianza de la clase k

N_k muestras de la clase k



Covarianza Intra-Clase (within-class covariance)

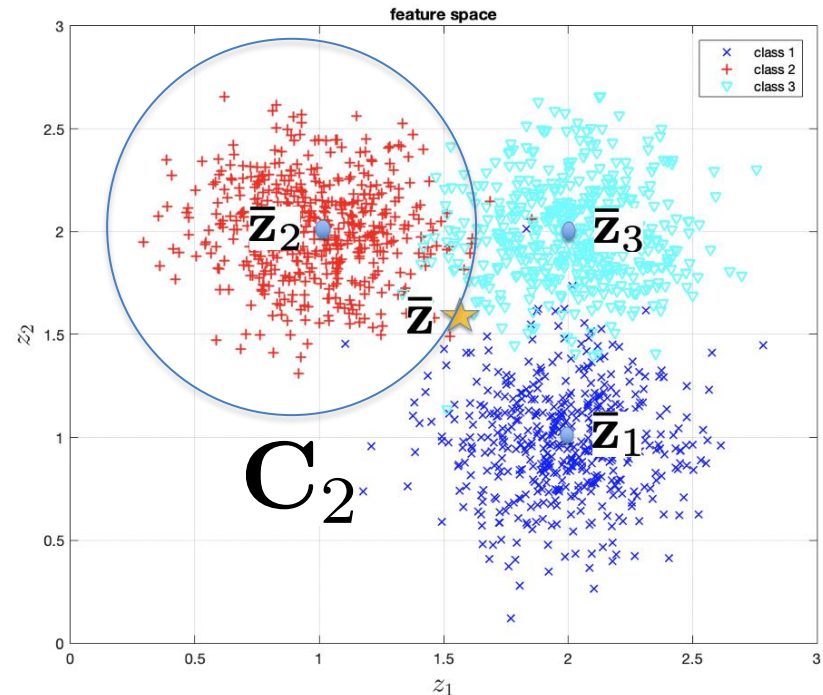
$$\mathbf{C}_w = \sum_{k=1}^K p_k \mathbf{C}_k,$$

Promedio ponderado de las Covarianzas de la cada clase

$$\mathbf{C}_k = \frac{1}{N_k - 1} \sum_{j=1}^{N_k} (\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)(\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)^T$$

Covarianza de la clase k

N_k muestras de la clase k



Covarianza Intra-Clase (within-class covariance)

$$\mathbf{C}_w = \sum_{k=1}^K p_k \mathbf{C}_k,$$

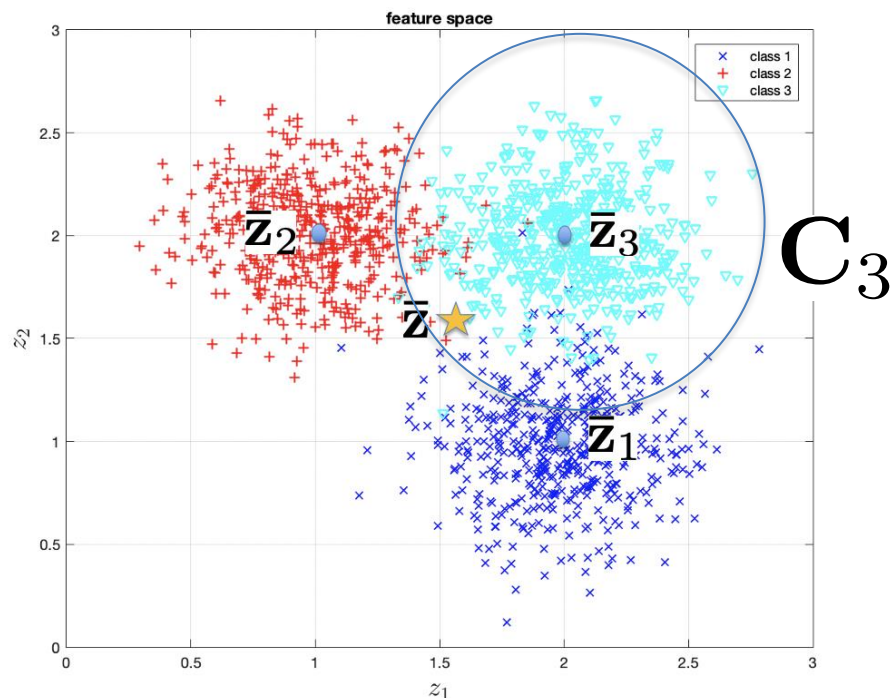
Promedio ponderado de las Covarianzas de la cada clase

$$\mathbf{C}_k = \frac{1}{N_k - 1} \sum_{j=1}^{N_k} (\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)(\mathbf{z}_{kj} - \bar{\mathbf{z}}_k)^T$$

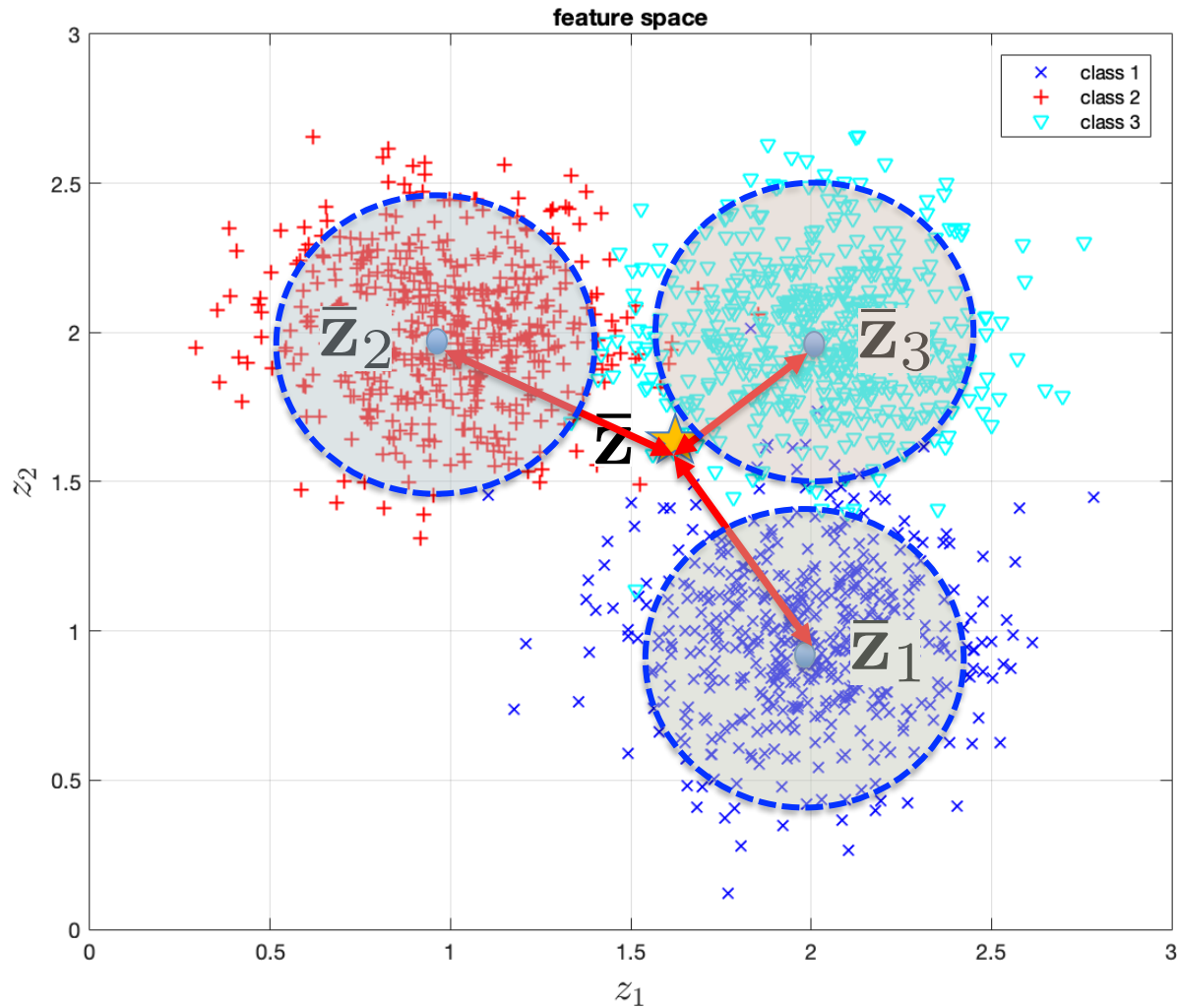


Covarianza de la clase k

N_k muestras de la clase k



Strategy: **Inter-Class HIGH** + **Intra-Class LOW**



Criterio de Fisher:

Covarianza Intra-Clase

$$\mathbf{C}_w = \sum_{k=1}^K p_k \mathbf{C}_k,$$

DEBE SER BAJA!

Covarianza Inter-Clase

$$\mathbf{C}_b = \sum_k p_k (\bar{\mathbf{z}}_k - \bar{\mathbf{z}})(\bar{\mathbf{z}}_k - \bar{\mathbf{z}})^\top$$

DEBE SER ALTA!

$$J = \text{spur} (\mathbf{C}_w^{-1} \mathbf{C}_b)$$



Suma de la diagonal

Criterio de Fisher:

Se puede aplicar para un problema de clasificación de K clases y m características.

Las matrices de covarianza son de $m \times m$, J es un escalar

$$J = \text{spur} \left(\mathbf{C}_w^{-1} \mathbf{C}_b \right)$$



Suma de la diagonal